Bringing Data to Action

AI at NASA: From Data to Insights To Actionable Intelligence

Kirk Borne

@KirkDBorne

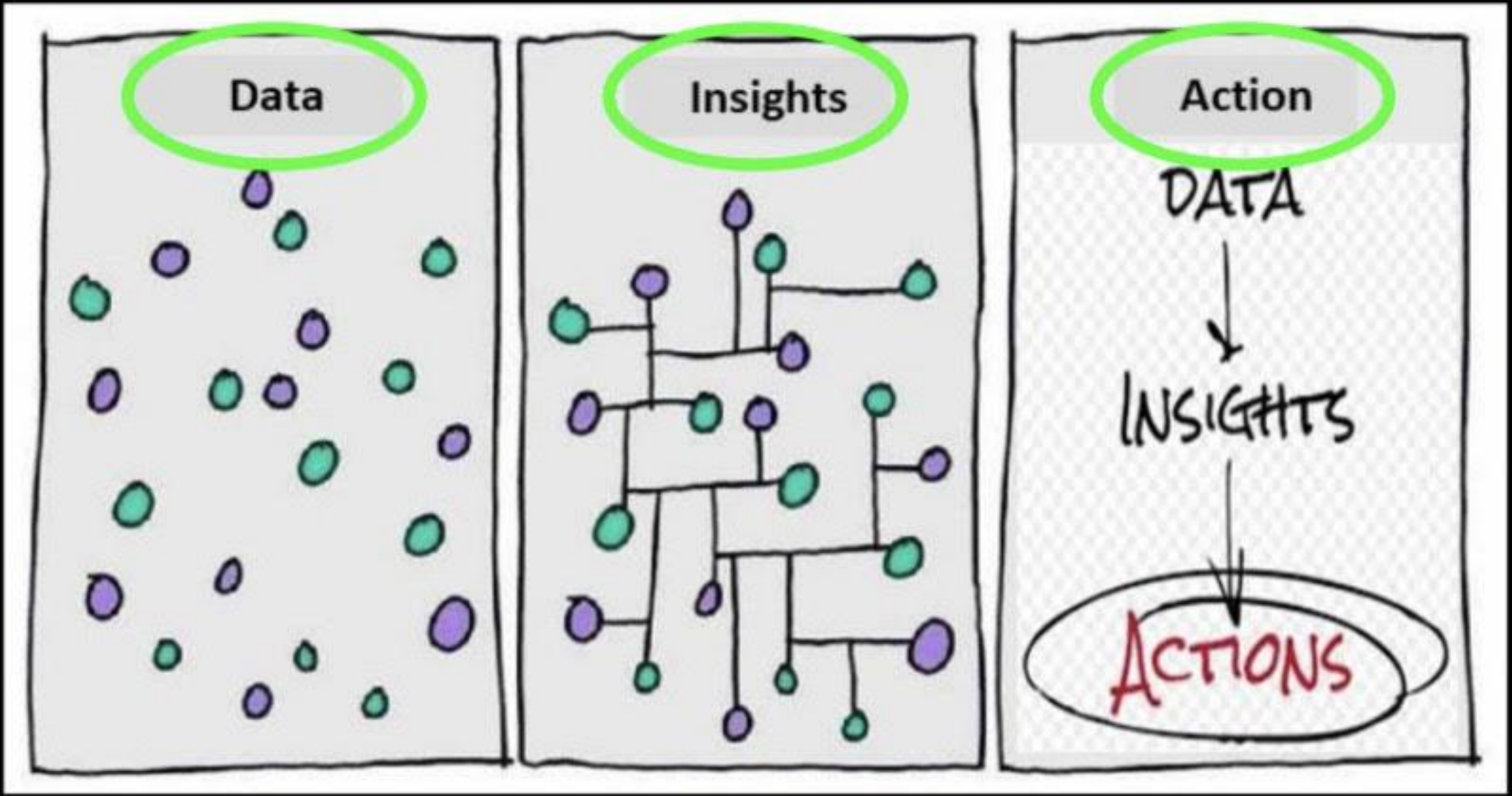
Booz Allen

Principal Data Scientist, Booz Allen Hamilton

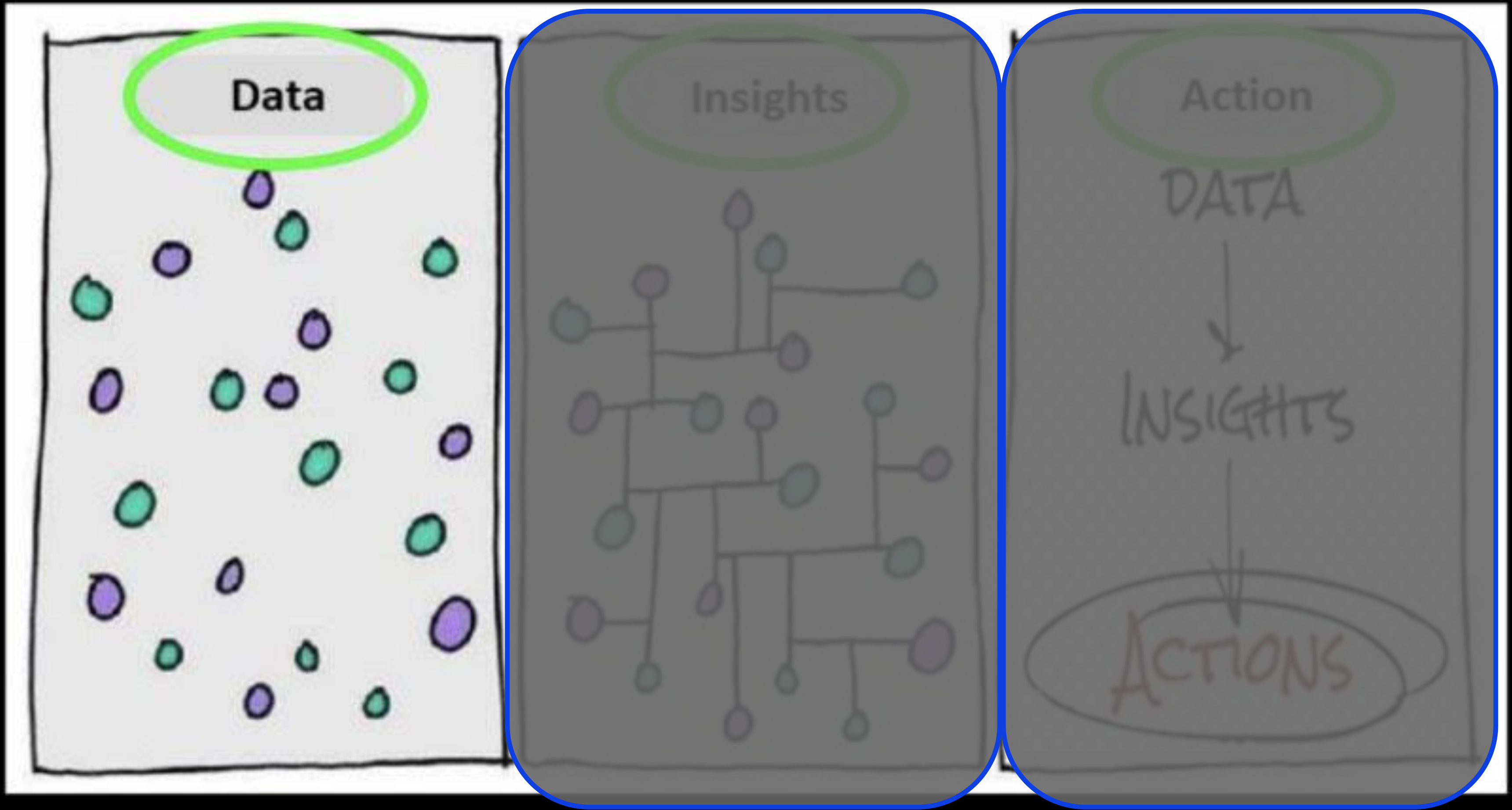
http://www.boozallen.com/datascience

From Data to Insights to Actionable Intelligence

From Data to Insights to Actionable Intelligence



From Data to Insights to Actionable Intelligence



DATA: Empowering Discovery in Science

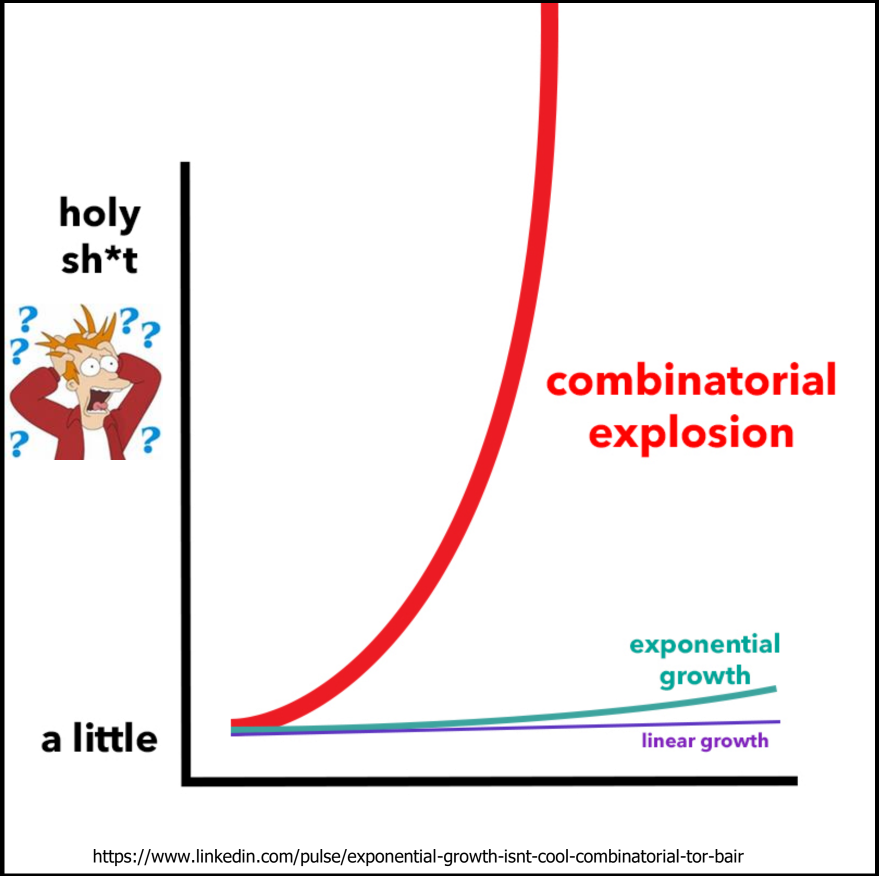
Source for graphic: https://goo.gl/PMHPJy

Ever since we first explored our world…

…we have asked questions about everything around us.

https://www.pinterest.com/pin/248683210647831264/

So, we have collected evidence (data) to answer our questions, which leads to more questions, which leads to more data collection, which leads to more questions, which leads to **BIG DATA!**



Knowledge is about connecting the dots.

@KirkDBorne

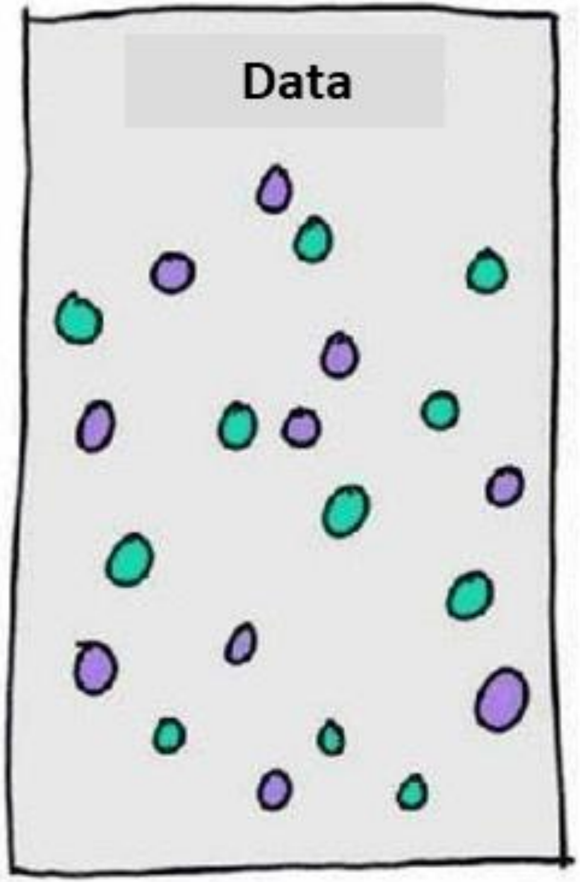
$y \sim x! \approx x \wedge x$

→ Combinatorial Growth!
(all possible interconnections, linkages, and interactions)

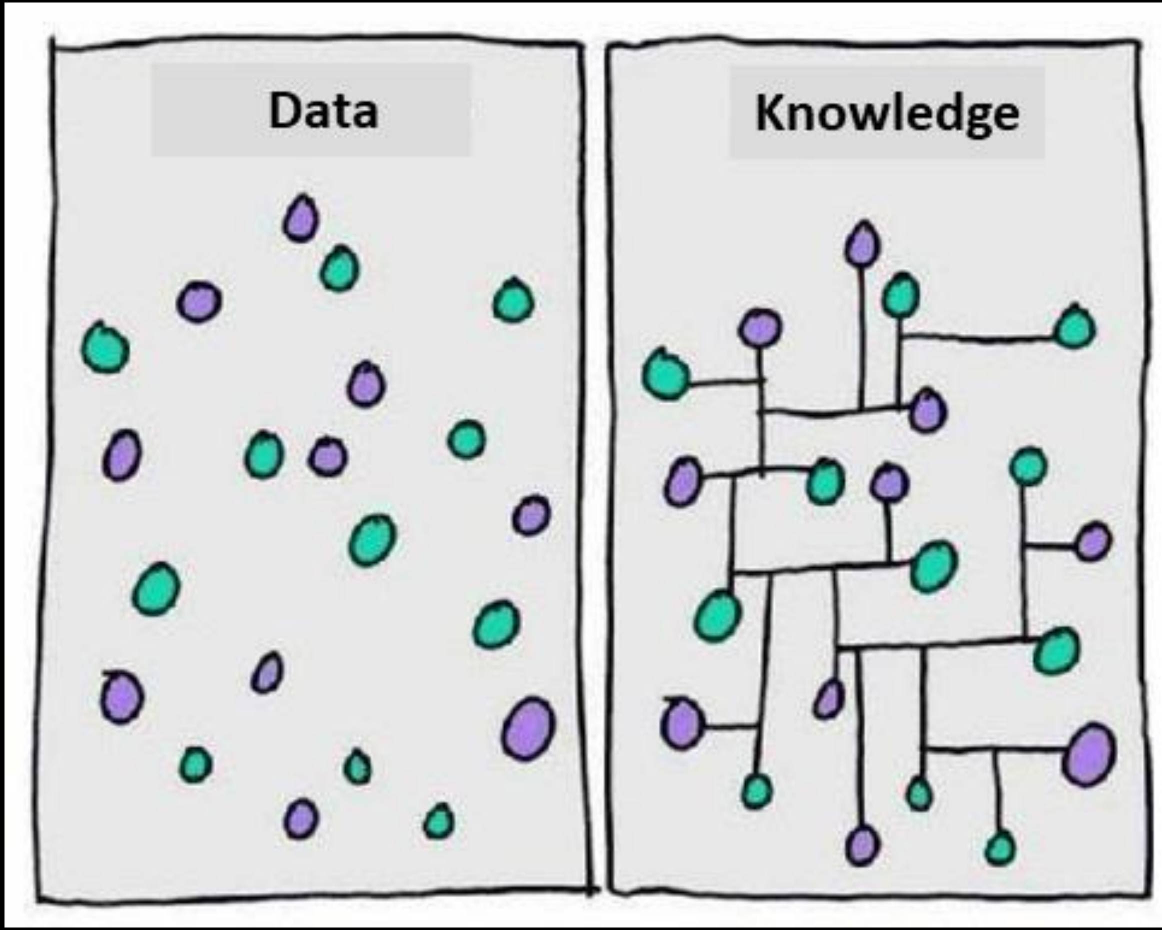
$y \sim 2 \wedge x$ (exponential growth)

$y \sim 2 * x$ (linear growth)

Data



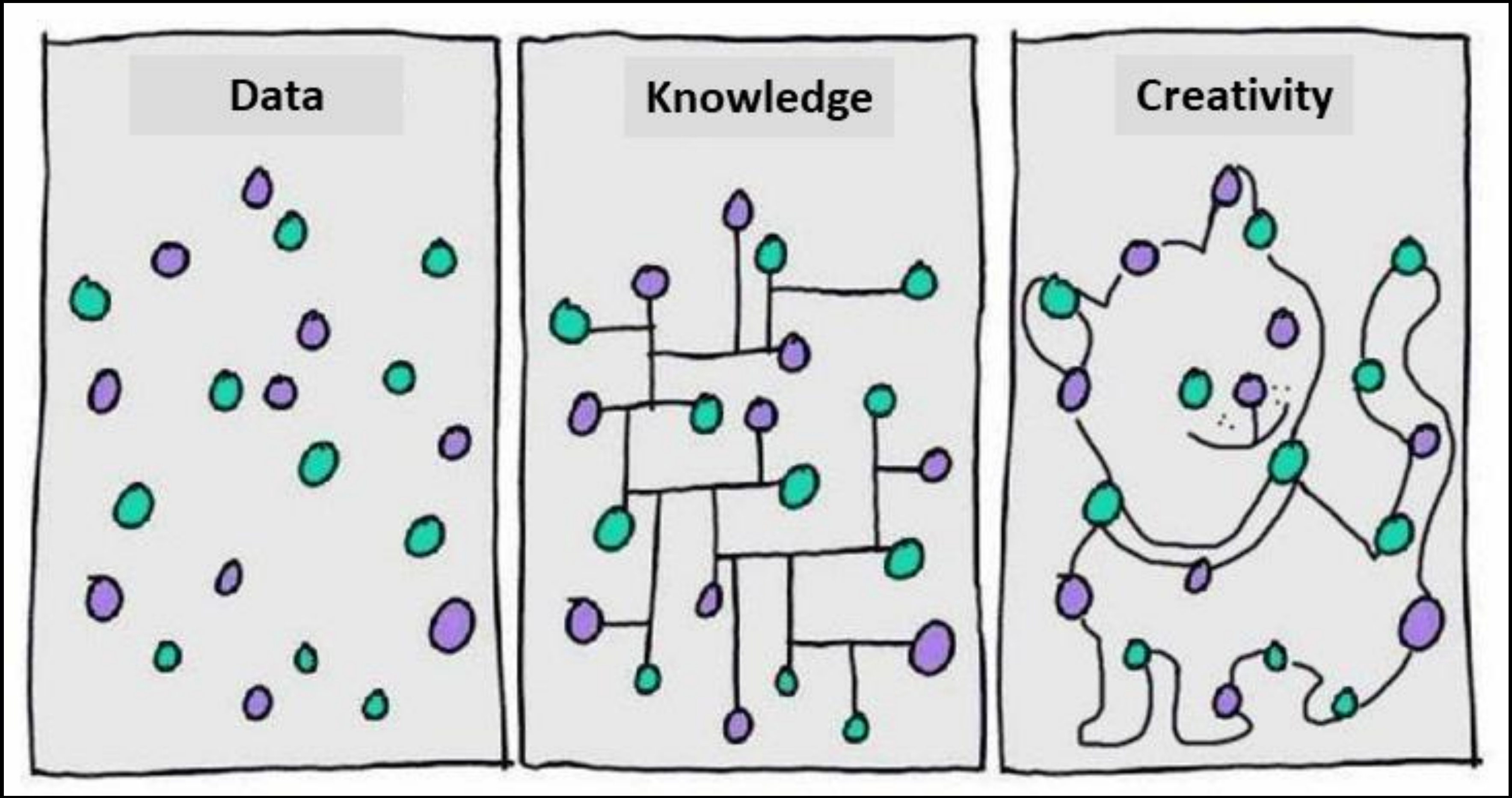
Data **Data Science (KDD)**



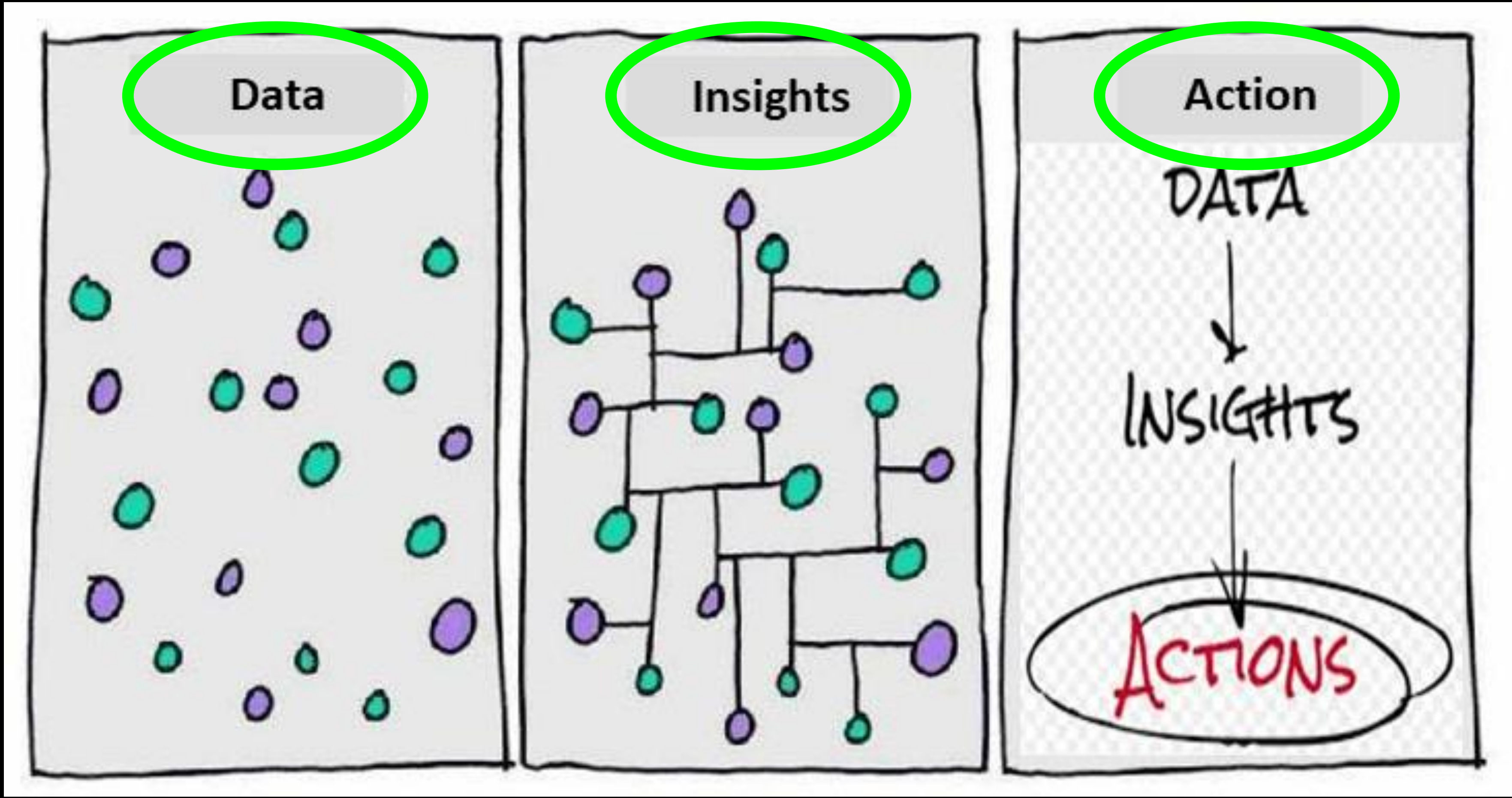
Knowledge is about connecting the dots.

@KirkDBorne

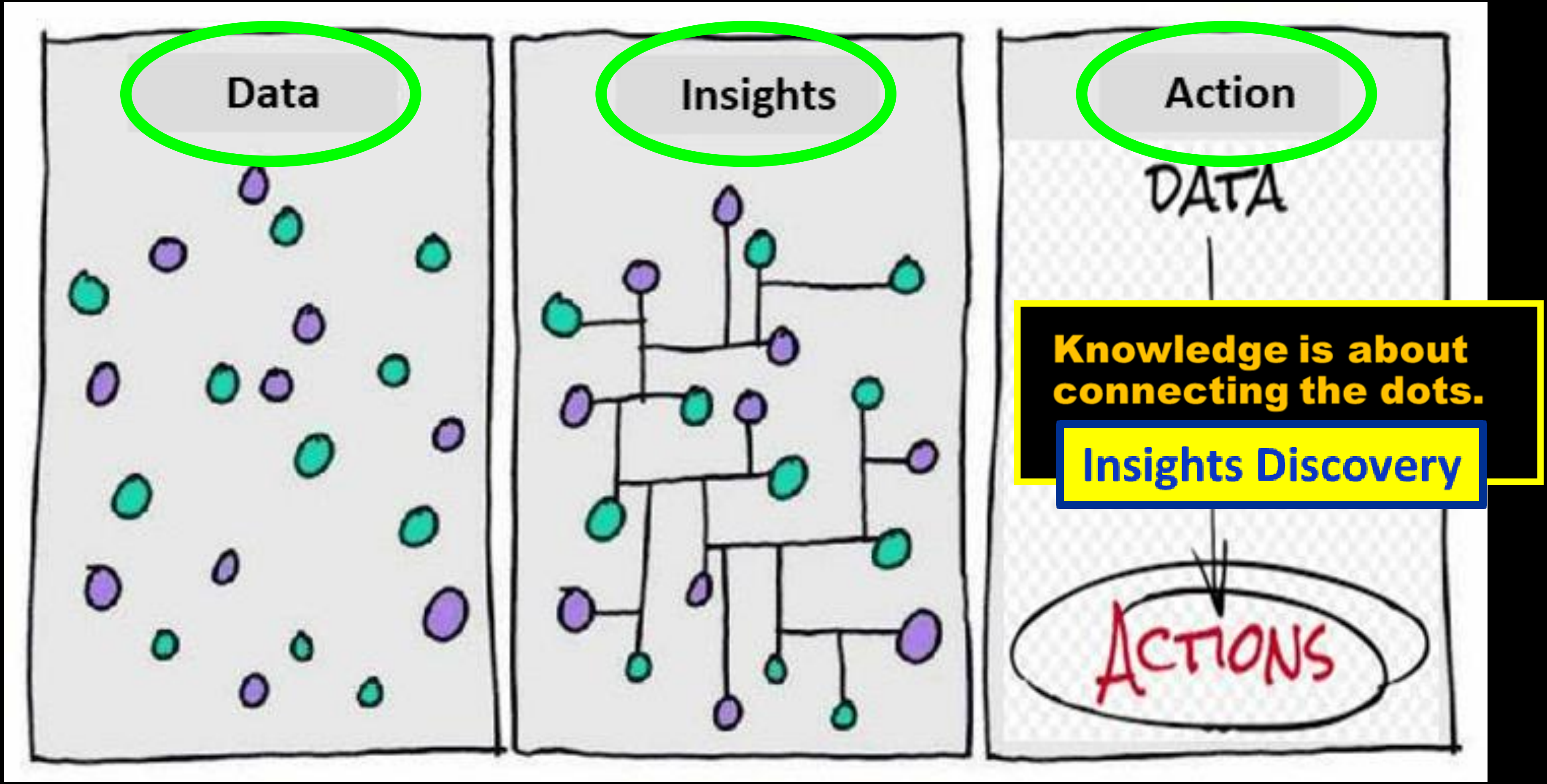
Data Data Science **Engineering Design**



Data Data Science (KDD) **A. I.**



Data Data Science (KDD) **A. I.**



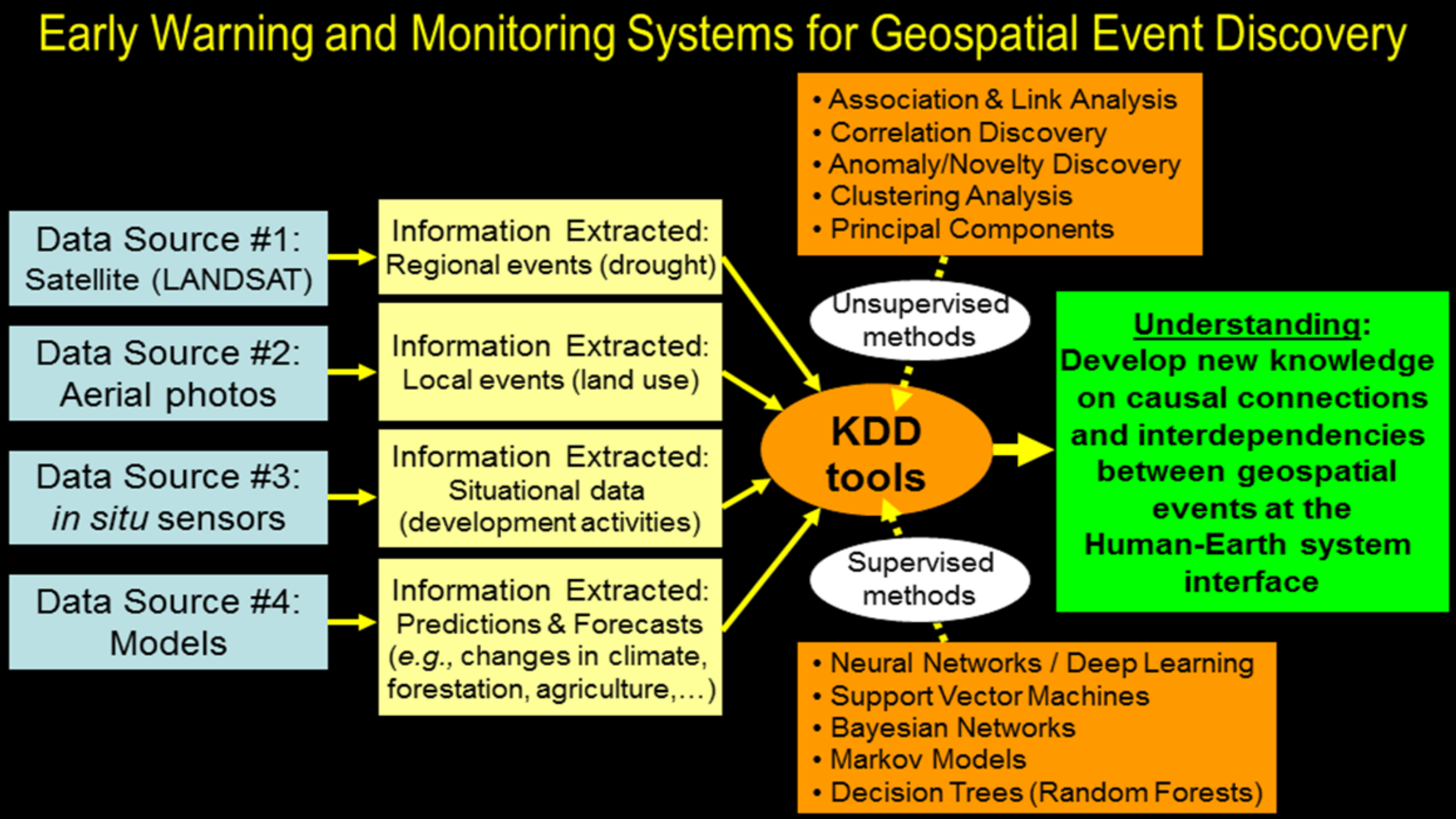
“Learn how to see. Realize that everything connects to everything else.”

— Leonardo da Vinci

...That’s cognitive!

GEOSPATIAL DATA SCIENCE USE CASE: ENVIRONMENTAL SCIENCE

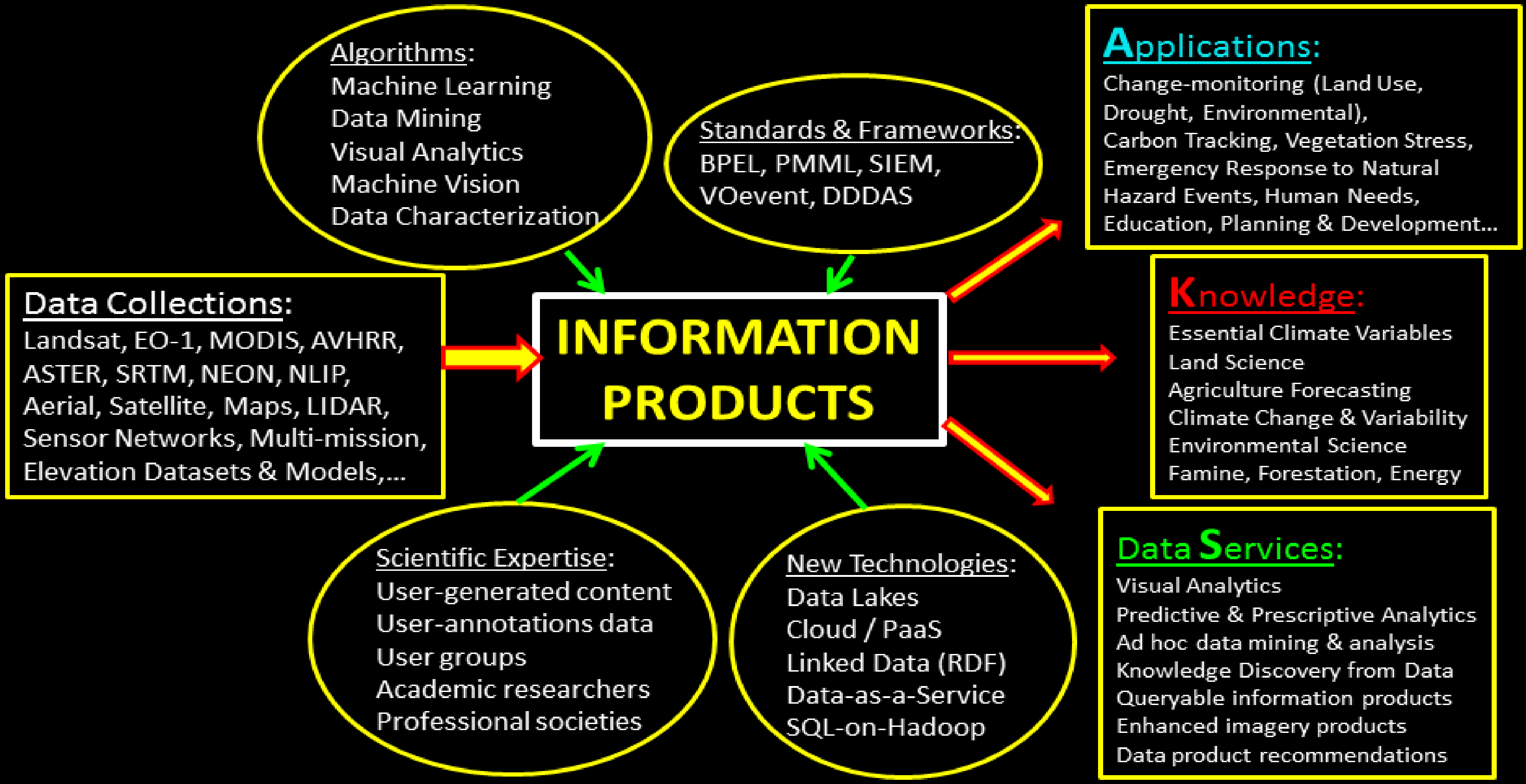
From Data to Information to Knowledge to Understanding



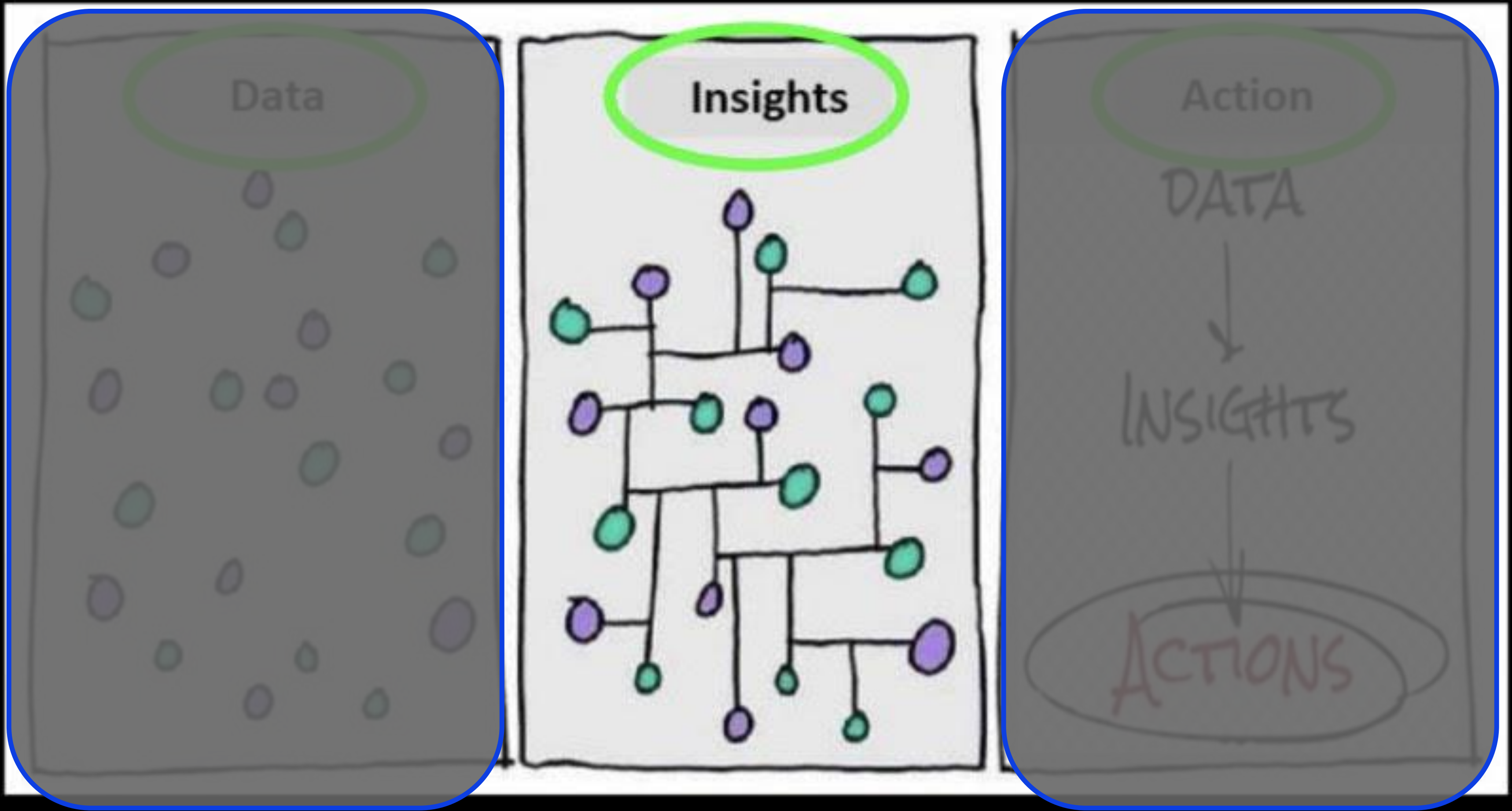
GEOSPATIAL DATA SCIENCE USE CASE: ENVIRONMENTAL SCIENCE

Data repositories & services for search, re-use, & building the knowledge graph!

The ASK pipeline = Applications, Services, Knowledge delivery from your data!



From Data to Insights to Actionable Intelligence



INSIGHTS: Learning Through Experience

Source for graphic: https://goo.gl/JznvrX

How do we Humans Learn?

Learning from sources of knowledge happens in two main ways:

DIRECT	INDIRECT
Facts and specific details that you retain in various methods…	*Experiences you must have on your own to retain…*
▪ Washington DC is the capital of the US	▪ Balancing on a bicycle
▪ More than half of the coastline of the entire United States is in Alaska	▪ Pronunciation of a foreign language

Moving between deductive and inductive reasoning during the learning cycle is a learning technique used by humans and machines

Source: Booz Allen Hamilton – https://twitter.com/KirkDBorne/status/1061286784024432640

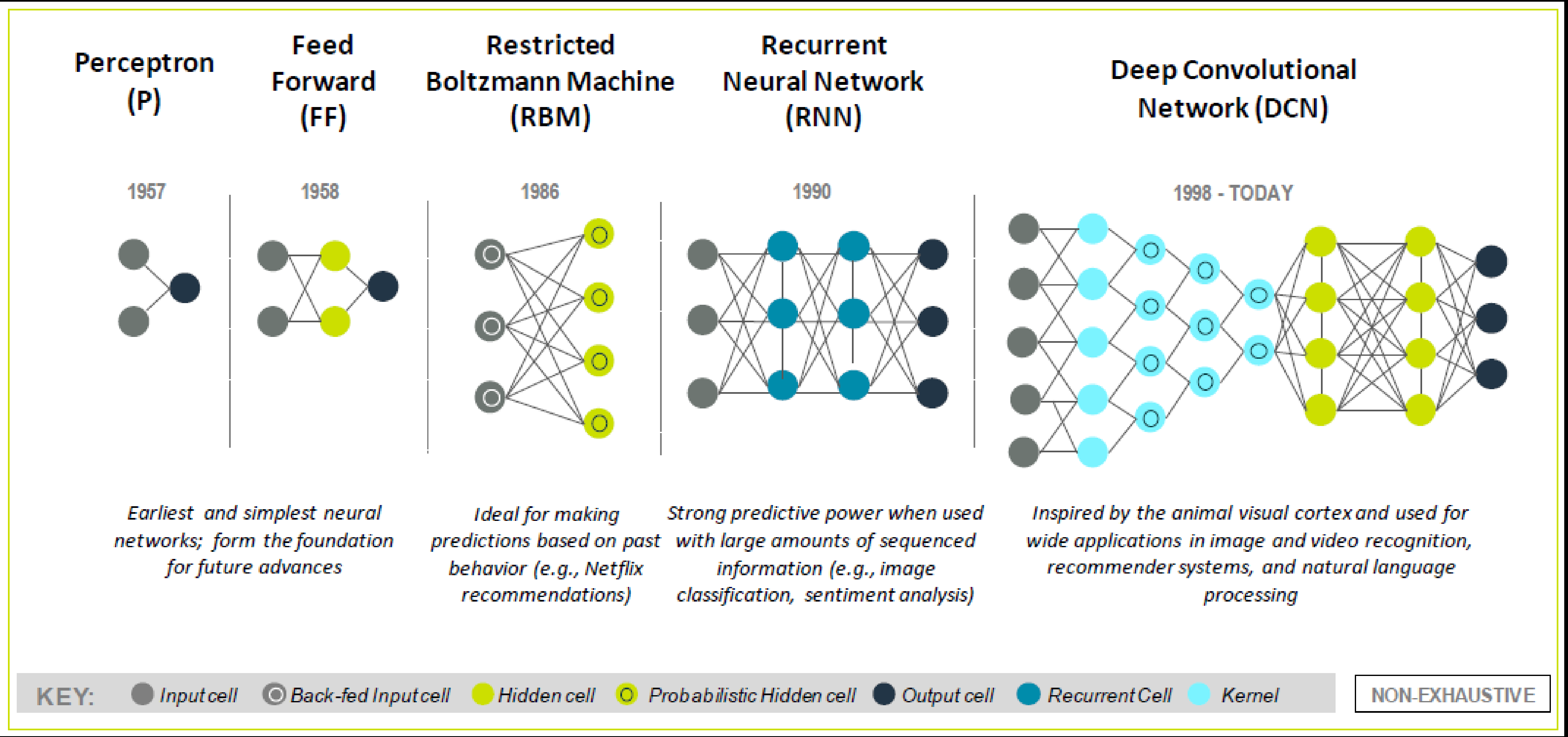
How do our Machines Learn?

Five approaches to structuring machine learning algorithms

	"TRIBE"	ORIGINS	MOTIVATION	TECHNICAL APPROACH
1. Fill in gaps in existing knowledge	**SYMBOLISTS**	Logic, Philosophy	Automate the scientific method	Inverse Deduction
2. Emulate the human brain	**CONNECTIONISTS**	Neuroscience	Reverse engineer the human brain via math model of neurons	Backpropagation
3. Simulate evolution over generations	**EVOLUTIONARIES**	Evolutionary Biology	Replicate the evolution of the human brain over generations	Genetic Programming
4. Systematically reduce uncertainty	**BAYESIANS**	Statistics	Test hypotheses to determine the certainty of knowledge	Probabilistic Inference
5. Find similarities between old and new	**ANALOGIZERS**	Psychology	Use previous problems / solutions and extrapolate into new context	Kernel Machines

Source: Booz Allen Hamilton – https://twitter.com/KirkDBorne/status/1061286784024432640

Increasingly sophisticated algorithms are powering today's advances in A.I.



Source: Booz Allen Hamilton – https://twitter.com/KirkDBorne/status/1061286784024432640

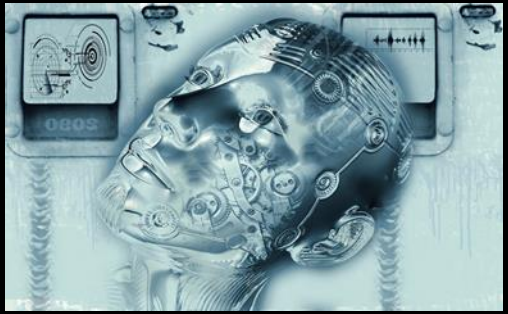
A.I. use cases and opportunities are ubiquitous (human-machine collaboration)

AI CAPABILITY	TECHNOLOGIES	EXAMPLE USE CASES	EXAMPLE APPLICATION
Pattern Recognition & Response *Maturing/Pilots and some scaled Deployment*	**Machine Learning Software and Platforms**	• Complex task automation • Real-time data analysis and response	*A cyber security algorithm detects, classifies, and prevents a network-based attack*
	Computer Vision	• Image/video tagging • Real-time video analysis • Sentiment analysis • Facial recognition • Scene analysis • Biometrics	*A video sensor on a drone identifies damage to an airfield runway*
	Natural Language Understanding	• Virtual assistants • Chatbots • Machine translation • Speech recognition • Language detection • Sentiment analysis • Text analysis • Report generation • Insight summarization	*Virtual assistants engage with citizens to ask about available camp grounds on recreation.gov*
	Autonomous Vehicles and Robotics	• Co-bots • Smart manufacturing • Smart logistics • Companion robots • Partially autonomous vehicles/unmanned systems	*A robotic surgeon performs surgery, automatically responding to changes in a patient's condition in real time*
Contextual Reasoning *In the lab*	**Semantic or "Cognitive" computing**	• Execution of tasks requiring context, judgment • Fully autonomous vehicles	*A vehicle drives down a crowded city road, responding to bad weather, unexpected pedestrian behavior, and obstacles in traffic*

Source: Booz Allen Hamilton – https://twitter.com/KirkDBorne/status/1061286784024432640

Some Quick Definitions:

"The 2 most important things in Data Science are the Data and the Science!"



- **Statistics** = the practice (and science) of collecting and analyzing numerical data.
- **Machine Learning (ML)** = mathematical **algorithms that learn from experience** (= pattern recognition from previous evidence).
- **Data Mining** = application of ML algorithms to data.
- **Artificial Intelligence (AI)** = application of ML algorithms to robotics and machines = taking actions based on data (**#bots**).
- **Data Science** = application of scientific method to discovery from data (including statistics, ML, and more: visual analytics, machine vision, computational modeling, semantics, graphs, network analysis, NLU, data indexing schemes [Google!], …).
- **Analytics** = the products of machine learning & data science.

The Real Power of A.I. – there is nothing "artificial" about it!

The New AI is better than Artificial Intelligence

Accelerated

Applied

Actionable

Assisted

Intelligence

Adaptable

Augmented

Amplified

Awesome

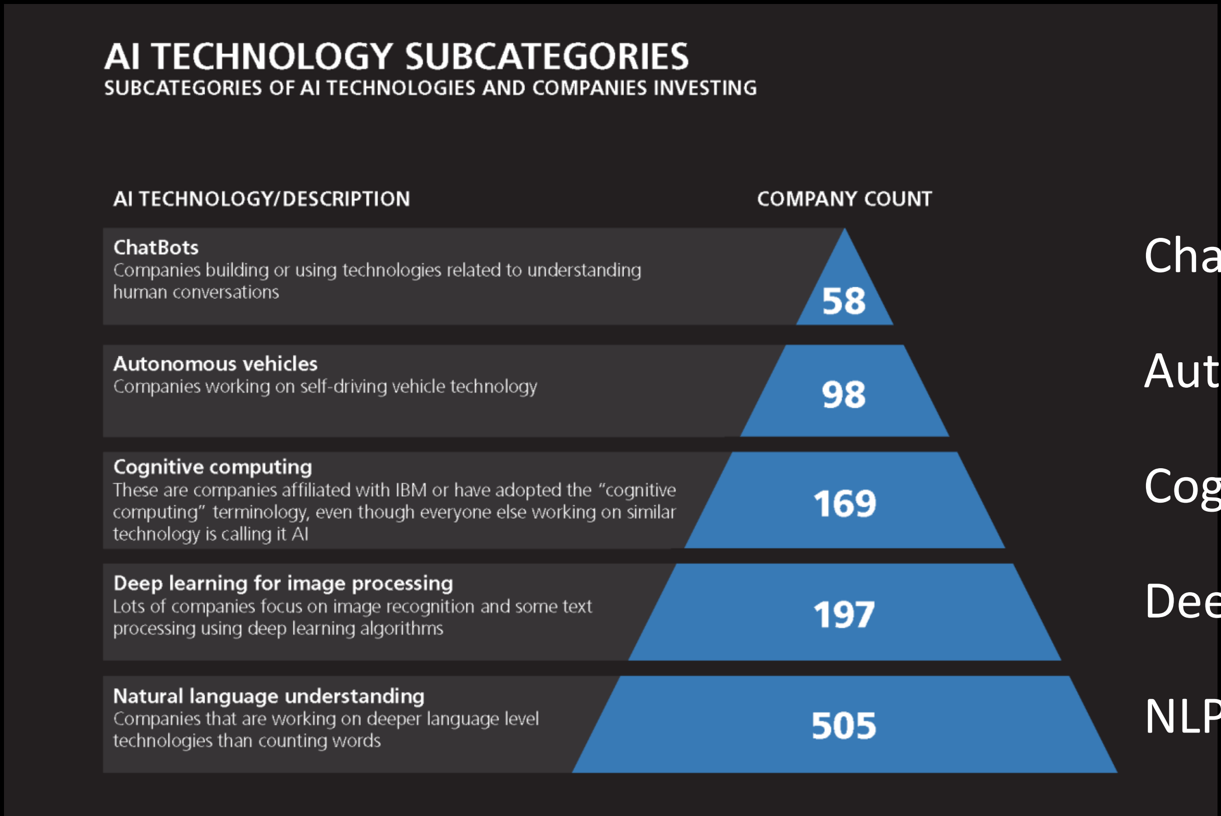
https://datamakespossible.westerndigital.com/real-power-ai/

Booz | Allen | Hamilton

@KirkDBorne

A.T. Tech Companies (the numbers in 2016)

“Pattern recognition is the basis of human intelligence." – Ray Kurzweil



https://www.oreilly.com/data/free/files/the-new-artificial-intelligence-market.pdf

Chatbots
(Conversational AI)
Autonomous Autos
(Self-driving “anything”)
Cognitive Computing
(Seeing & Deciding like us)
Deep Learning
(CV: Computer Vision)
NLP = NLU + NLG
(Narrative Science)

The Rich Landscape of AI Startups in 2018

https://www.cbinsights.com/research/artificial-intelligence-top-startups/

An "Easy Button" for taking Data to Action through Machine Learning

- Pattern Discovery (Detection)
  - D2D: data-to-discovery
- Pattern Recognition
  - D2D: data-to-decisions
- Pattern Exploration
  - D2D: data-to-dollars (innovation)
- Pattern Exploitation
  - D2V: Data-to-Value (action)
  - D2A: Data-to-Action (value)

Bringing Data to Action

The Goal of Machine Learning

"…is to use algorithms to learn from data (training data = "experience"), in order to build generalizable models that give accurate classifications or predictions, or to find (useful) patterns, particularly in new and previously unseen data."

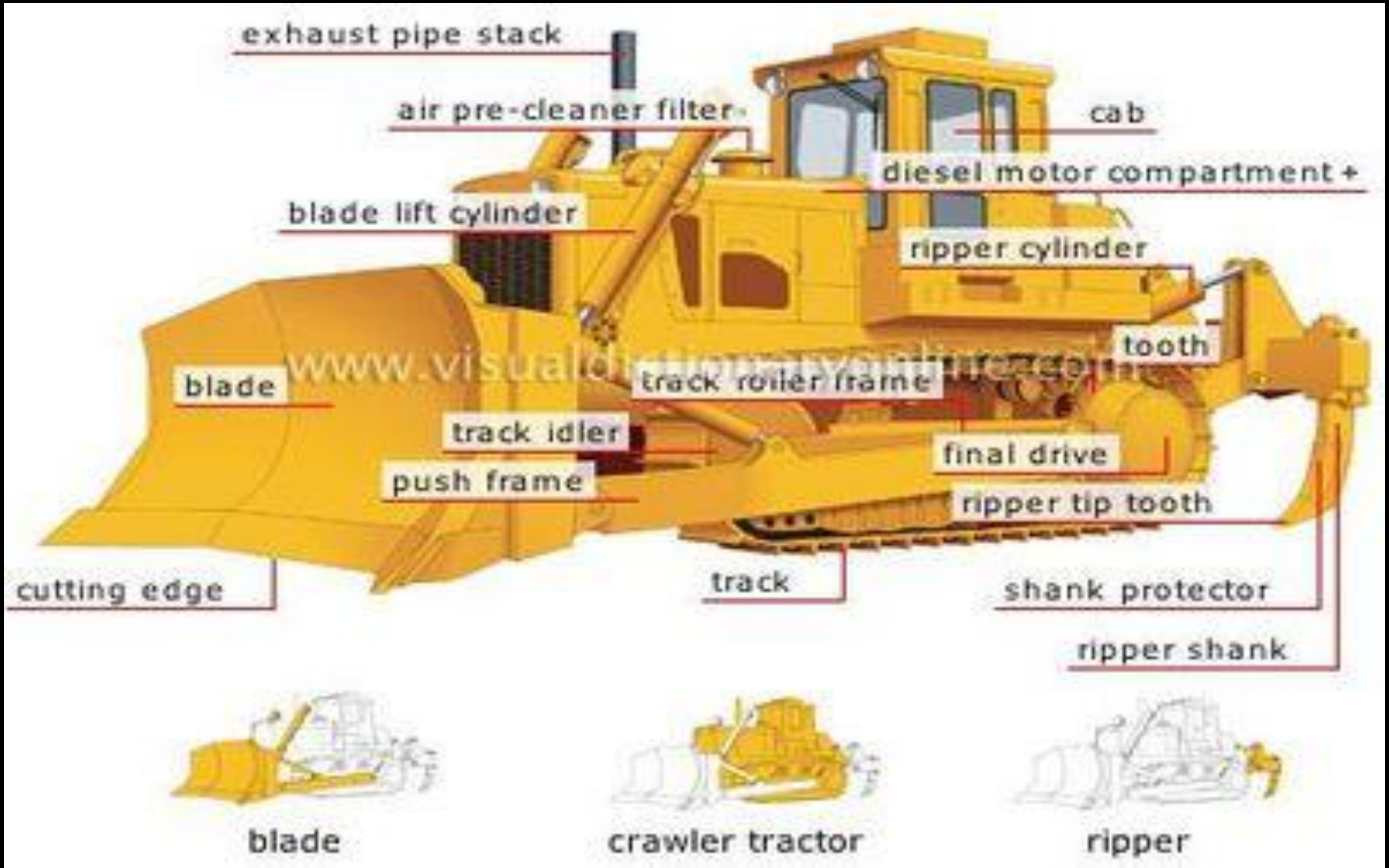
(the key is GENERALIZATION!)

https://www.innoarchitech.com/machine-learning-an-in-depth-non-technical-guide/

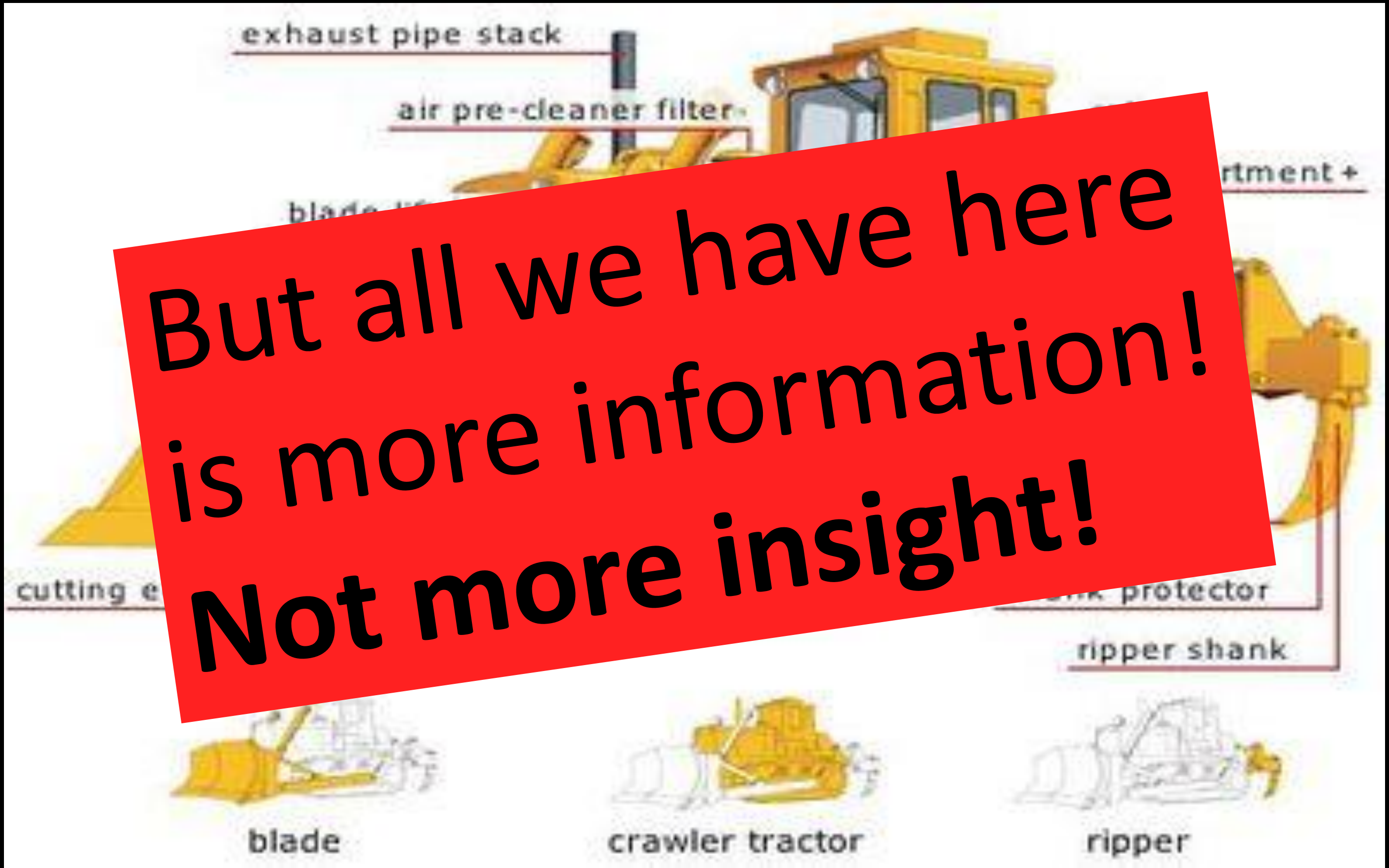
How does a Data Scientist build a model of a complex dynamic system?

We might start by modeling a complex system like this…

We can add more features to model the system with higher fidelity …



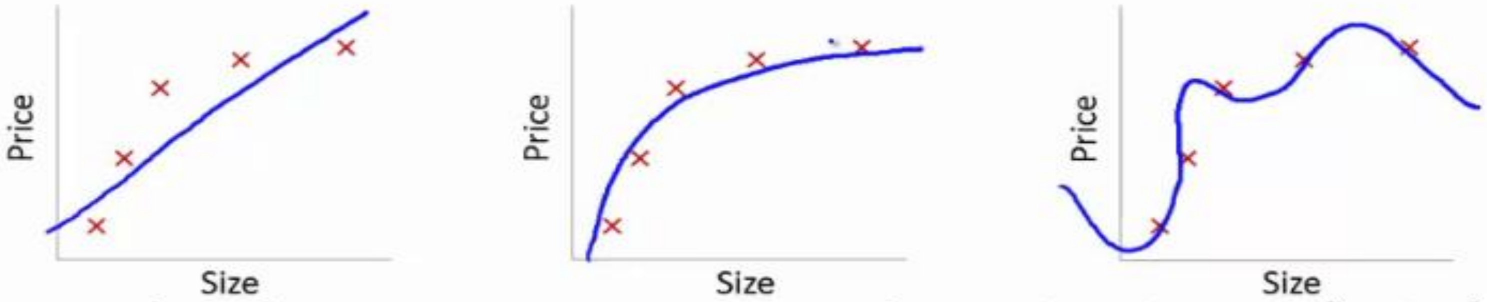
We can add more features to model the system with higher fidelity …



Pattern Discovery is easy, but Pattern Exploitation requires more data science…

Generalization is key!

Insights Discovery



$\theta_0 + \theta_1 x$ | $\theta_0 + \theta_1 x + \theta_2 x^2$ | $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High bias (underfit)

"Just right" **(The Goldilocks model)**

High variance (overfit)

The most generally useful model captures the fundamental pattern in the data and takes into account the natural variance in the data.

Source for graphic: http://www.holehouse.org/mlclass/10_Advice_for_applying_machine_learning.html

Generalization is key to gaining insight

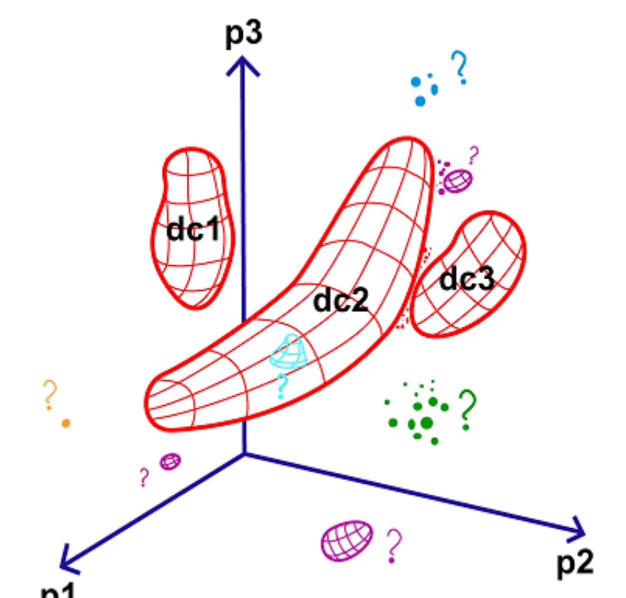
in·sight

/ˈinˌsīt/

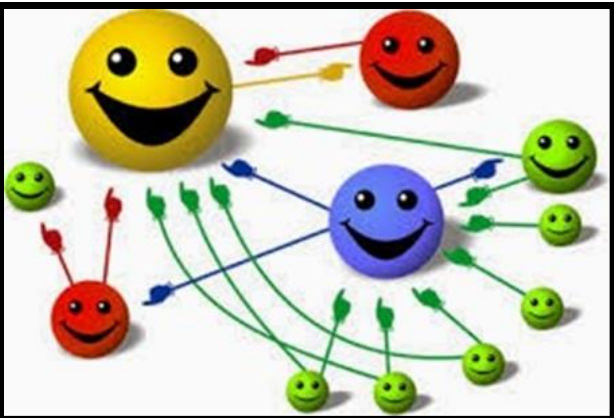
noun

[1] the capacity to gain an accurate and deep intuitive understanding of a person or thing.

Insights Discovery



(Graphic by S. G. Djorgovski, Caltech)



4 Types of Discovery from Data:

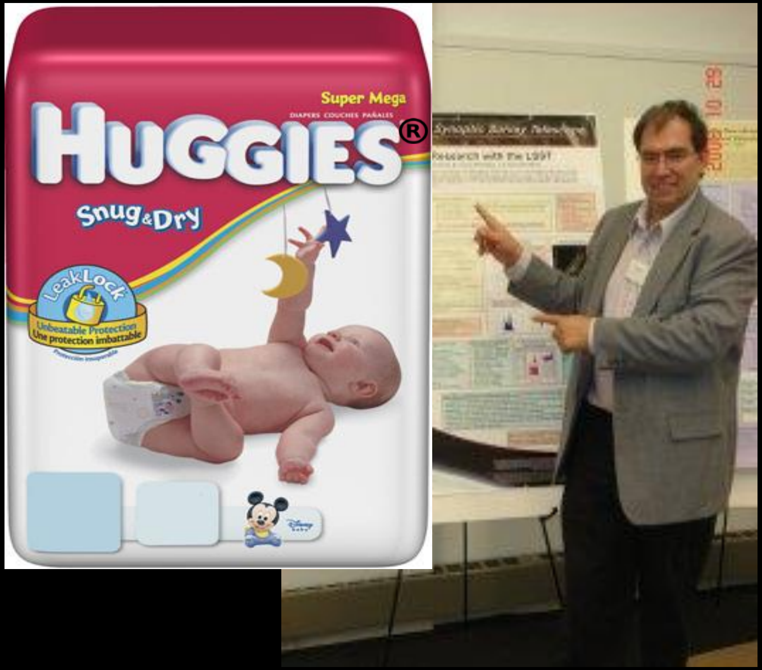
1) **Class Discovery:** Find the categories of objects (population segments), events, and behaviors in your data. + Learn the rules that constrain the class boundaries (that uniquely distinguish them).
2) **Correlation (Predictive and Prescriptive Power) Discovery:** (**INSIGHT DISCOVERY**) – Find trends, patterns, and dependencies in data that reveal the governing principles or behavioral patterns (the object's "DNA").
3) **Outlier / Anomaly / Novelty / Surprise Discovery:** Find the new, surprising, unexpected one-in-a-[million / billion / trillion] object, event, or behavior.
4) **Association (or Link) Discovery:** (Graph and Network Analytics) – Find both the typical (usual) and the atypical (unusual, interesting) data associations / links / connections in your domain.

2 Examples of Insights Discovery

1) **Associations**
2) Graphs

Association Discovery Example #1

- **Classic Textbook Example of Data Mining** (Legend?): Data mining of grocery store logs indicated that men who buy diapers also tend to buy beer at the same time.



Association Discovery Example #2

- **Amazon.com** mines its customers' purchase logs to recommend books to you: *"People who bought this book also bought this other one."*

Association Discovery Example #3

- **Netflix** mines its video rental history database to recommend rentals to you based upon other customers who rented similar movies as you.

Association Rule Discovery for Hurricane Intensification Forecasting

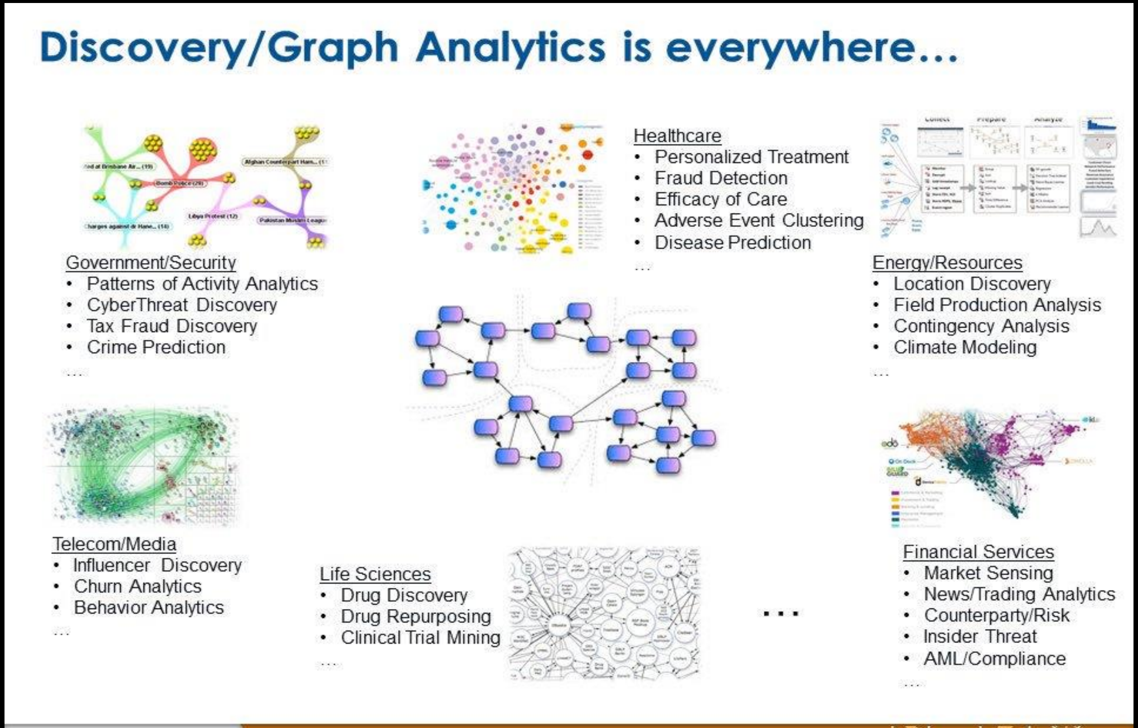
- Research by GMU geoscientists
- Predict the final strength of hurricane at landfall.
- Find co-occurrence of final hurricane strength with specific values of measured physical properties of the hurricane *<u>while it is still over the ocean</u>*.
- **Insights Discovery** : association rule discovery prediction is better than National Hurricane Center prediction!
- Research Paper by GMU scientists: https://ams.confex.com/ams/pdfpapers/84949.pdf

2 Examples of Insights Discovery

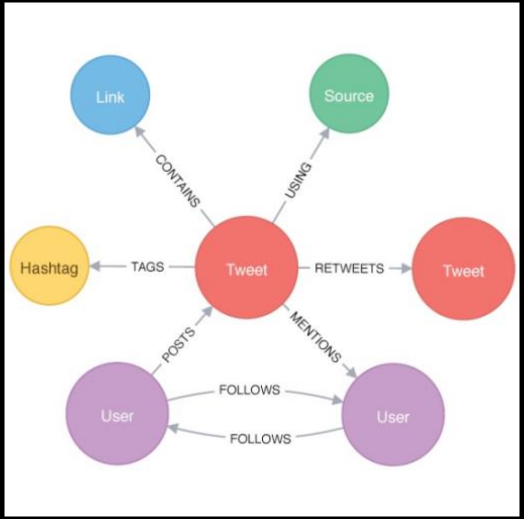
1) Associations
2) **Graphs**

"All the World is a Graph" - Shakespeare?

The natural data structure of the world is not rows and columns, but a Graph!

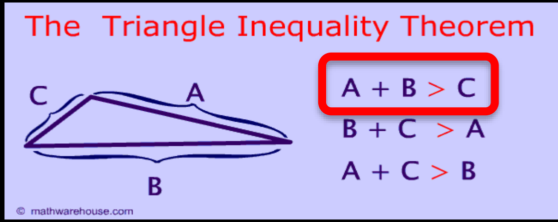


(Graphic by Cray, for Cray Graph Engine CGE)



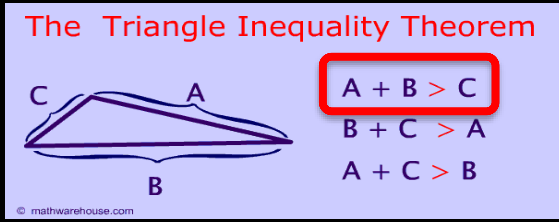
http://www.cray.com/products/analytics/cray-graph-engine

Simple Example of the Power of Graph: Semi-Metric Space

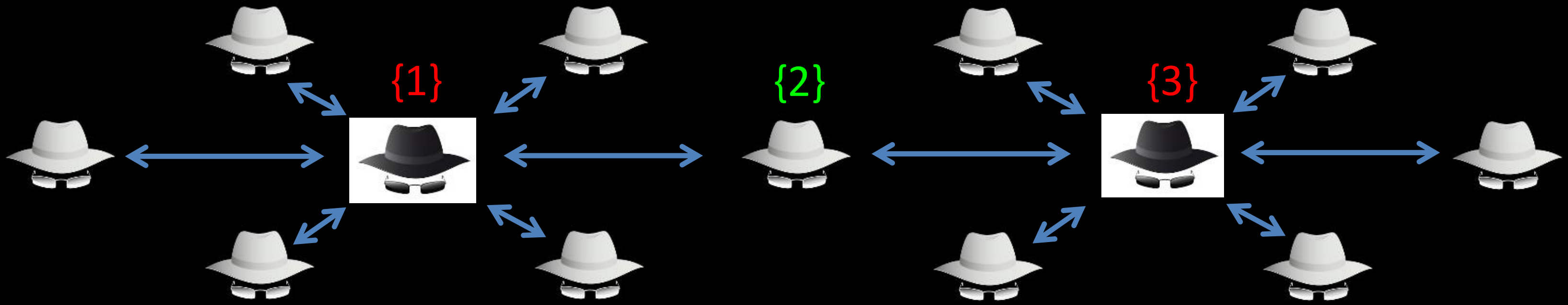


- Entity {1} is linked to Entity {2} (small distance A)
- Entity {2} is linked to Entity {3} (small distance B)
- Entity {1} is *not* linked directly to Entity {3} (Similarity Distance C = infinite)
- Similarity Distances between A, B, and C violate the triangle inequality!

Simple Example of the Power of Graph: Semi-Metric Space

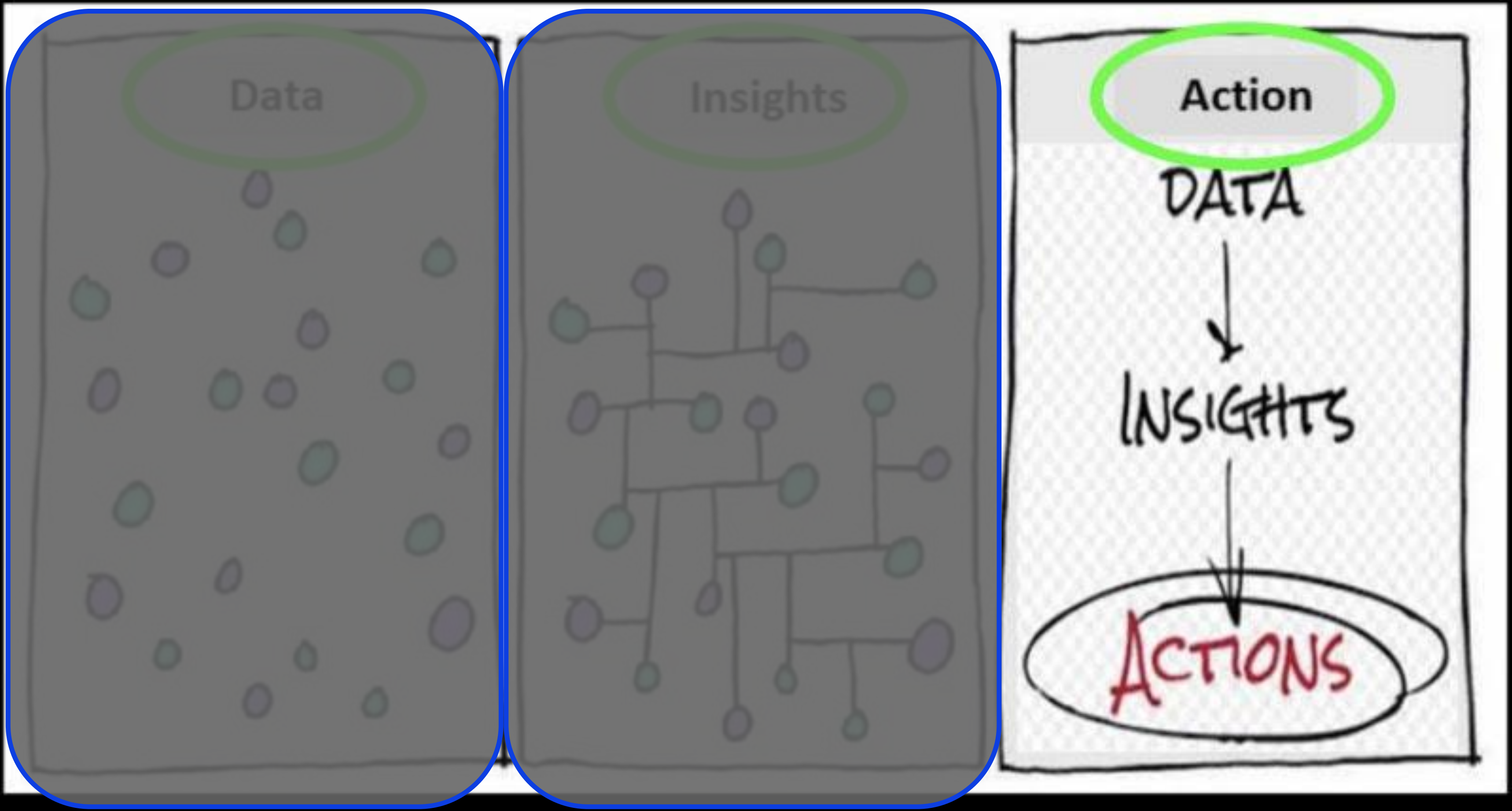


- Entity {1} is linked to Entity {2} (small distance A)
- Entity {2} is linked to Entity {3} (small distance B)
- Entity {1} is *not* linked directly to Entity {3} (Similarity Distance C = infinite)
- Similarity Distances between A, B, and C violate the triangle inequality!



- The connection between black hat entities {1} and {3} never appears explicitly within a transactional database.
- Examples of **Insights Discovery** from Graphs: (a) Research Discoveries across disconnected journals, through linked semantic assertions; (b) Journeys; (c) Safety Incident Causal Factor Analysis; (d) Marketing Attribution Analysis; (e) Discovery of Fraud networks, Illegal goods trafficking networks, Money-Laundering networks.

From Data to Insights to Actionable Intelligence

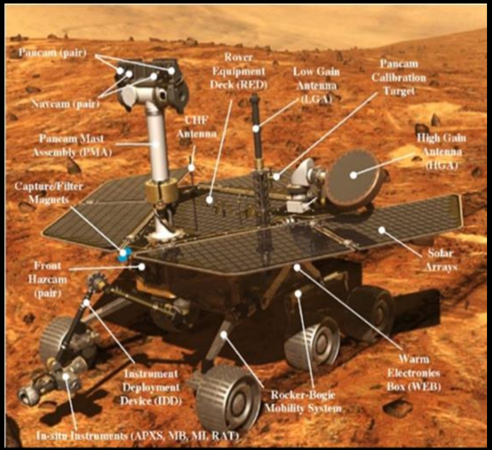


ACTIONABLE INTELLIGENCE:
Seeing Our Way Forward

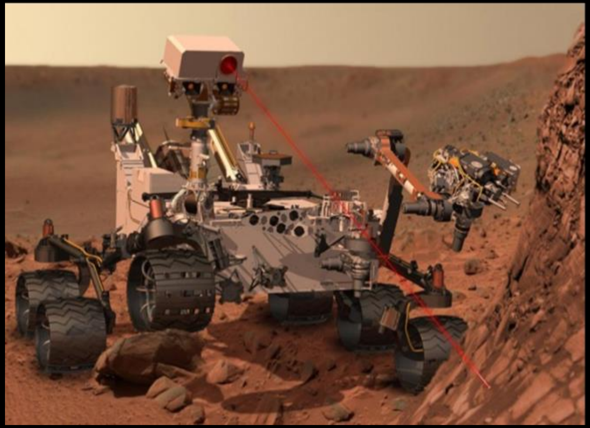
Source for graphic: https://goo.gl/ksfqib

Mars Rovers :

- Your Smart Data gatherer
- Actionable intelligent agent
- Autonomous decision system for the Data-informed Journey



http://bizbench.com/wisdom-pyramid/



Taking Data to Information to Knowledge to Understanding to Action

Taking Data to Information to Knowledge to Understanding to Action

Mars Rover:

1) Drive around surface of Mars – take samples of items (Mars rocks):

2) Perform Intelligent Data Operations (Analytics at the Edge):

- Supervised Learning
  - Search for items with known characteristics, and assign items to known classes
- Unsupervised Learning
  - Discover what types of items are present, without preconceived biases; find the set of unique classes of items; discover unusual associations; discover trends and new directions (new leads) to follow
- Semi-supervised Learning
  - Find the rare, one-of-kind, most interesting items, behaviors, contexts, or events
- Reinforcement Learning
  - Enact Intelligent Data Understanding & Decision Support at the point of data collection to maximize cumulative rewards => **Business Goal Maximization through Feedback:**
    - *"stay here and do more"*; or else *"follow trend to a more interesting location"*
    - *"send discoveries to human analysts immediately"*; or *"send informational results later"*

Smart Sensors & Sentinels for Data-Driven Sense-Making and Decision Support

From Sensors to Sentinels to Sense

(for any application domain with streaming data from sensors)

- New knowledge and insights are acquired by monitoring and mining actionable data from all digital inputs (**Sensors!**)

- Alerts are triggered **autonomously, without intervention** (if permitted), applying machine learning and actionable business decision rules for pattern detection and diagnosis. (**Sentinels!** = embedded machine learning / data science algorithms, at the point of data collection = trained to minimize False Positives and "Alarm Fatigue")

- "Smart Sensors" (powered by Machine Learning-enabled sentinels) will therefore deliver actionable intelligence (**Sense!**)

Dynamic Data-Driven Application Systems (DDDAS)

http://dddas.org

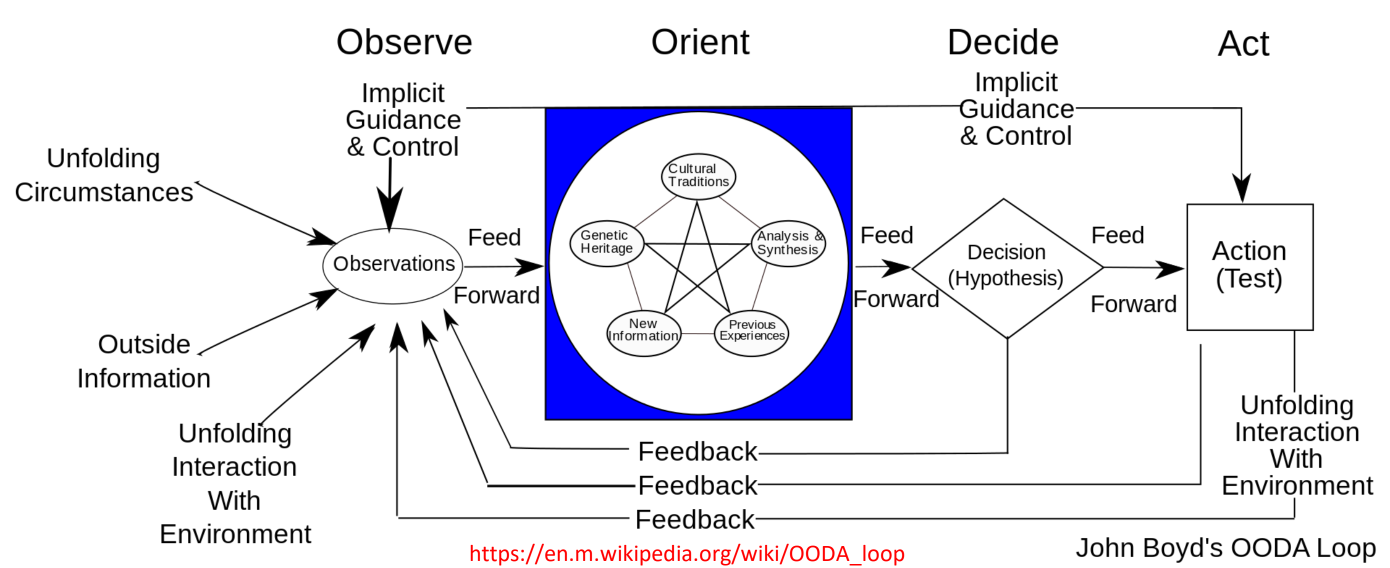
- **4 steps from data to action in DDDAS = MIPS**:
  - **M**easurement – **I**nference – **P**rediction – **S**teering
- The "**M**" in MIPS = Measurement (any type of sensor data):
  - Health & Epidemic monitoring systems, Web user interactions & actions (web analytics data), Cyber network usage logs, Customer Service interactions, Social network sentiment, Machine logs (of any kind), Manufacturing sensors, Financial transactions, National Security, Utilities and Energy, Satellite imagery, Remote Sensing, Tsunami warnings, Weather/Climate events, Astronomical sky events, world events, …
- Machine Learning enables the "**IP**" steps in MIPS: **Insights Discovery**
  - Pattern (Segment) Discovery
  - Correlation (Trend) Discovery
  - Novelty (Anomaly) Discovery
  - Association (Link) Discovery
- The "**S**" in MIPS = Steering = Action!
  - Robotic Process Automation (RPA)

A.I. :
- Gather actionable insights from streaming sensor data
- Automate any data-driven operational system

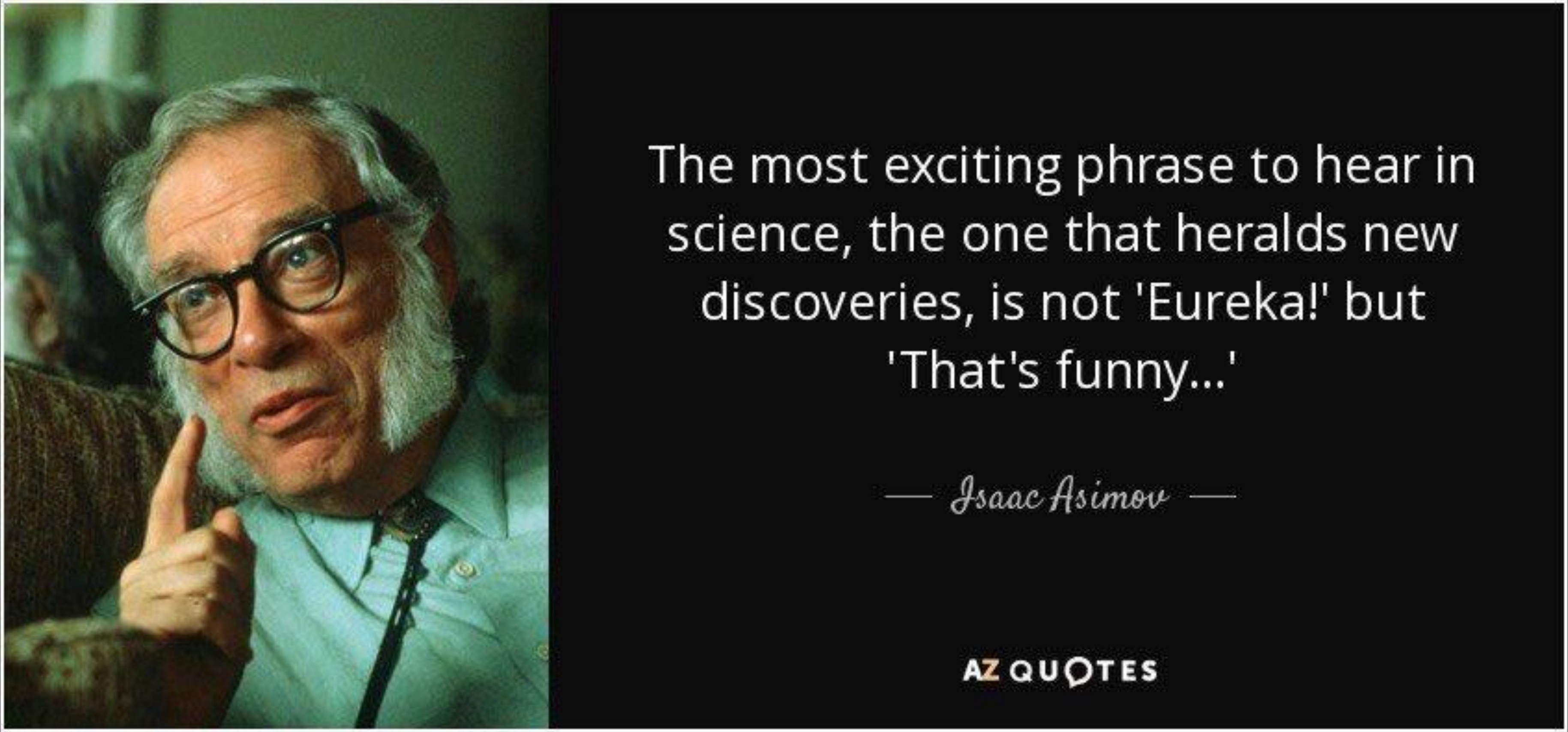
Dynamic Data-Driven Application Systems (DDDAS)

http://dddas.org

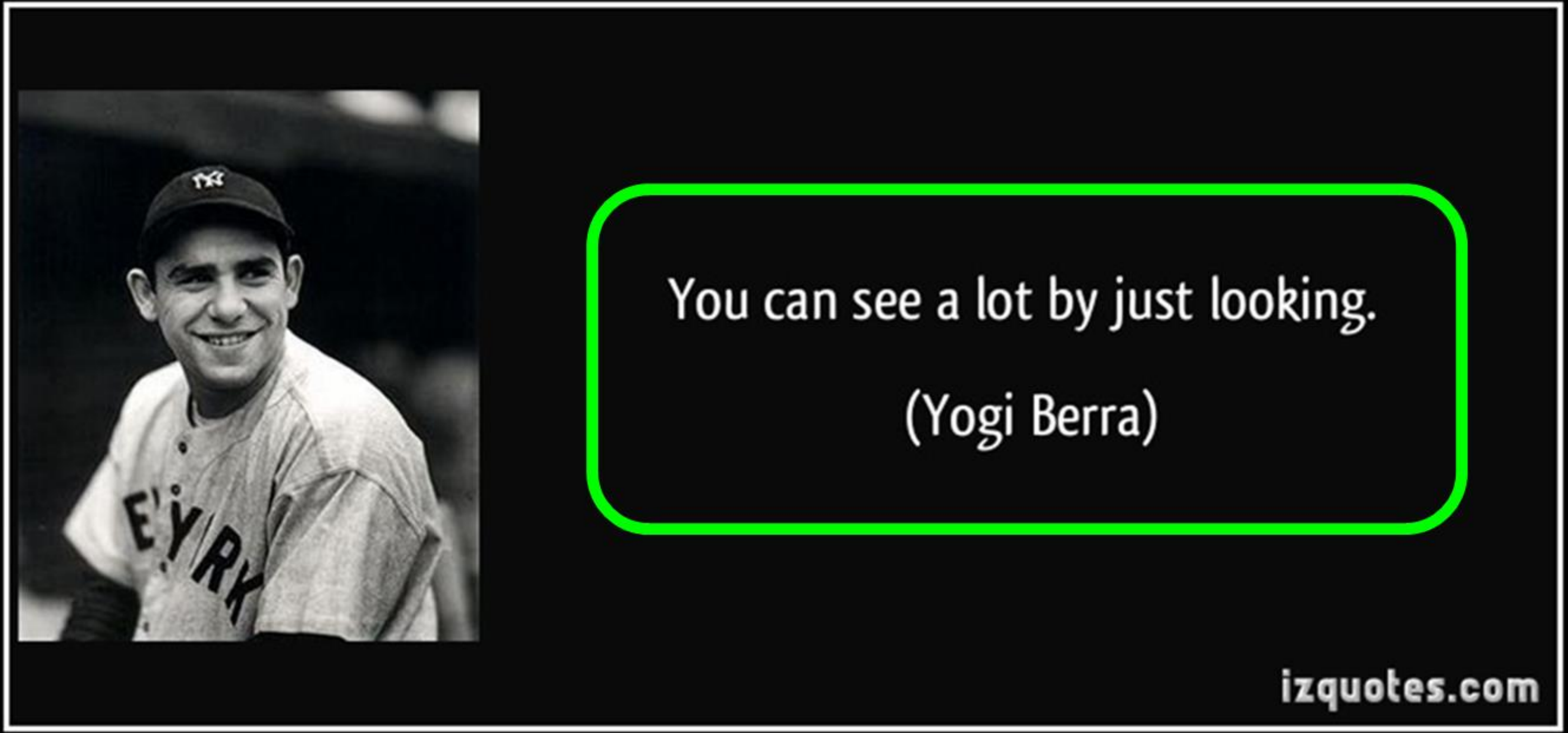
- **4 steps from data to action in DDDAS = MIPS:**
  - **M**easurement – **I**nference – **P**rediction – **S**teering
- This is essentially the same as the **OODA Loop** for military action and command decision-making (e.g., fighter pilots):
  - **O**bserve – **O**rient – **D**ecide – **A**ct **(OODA)**



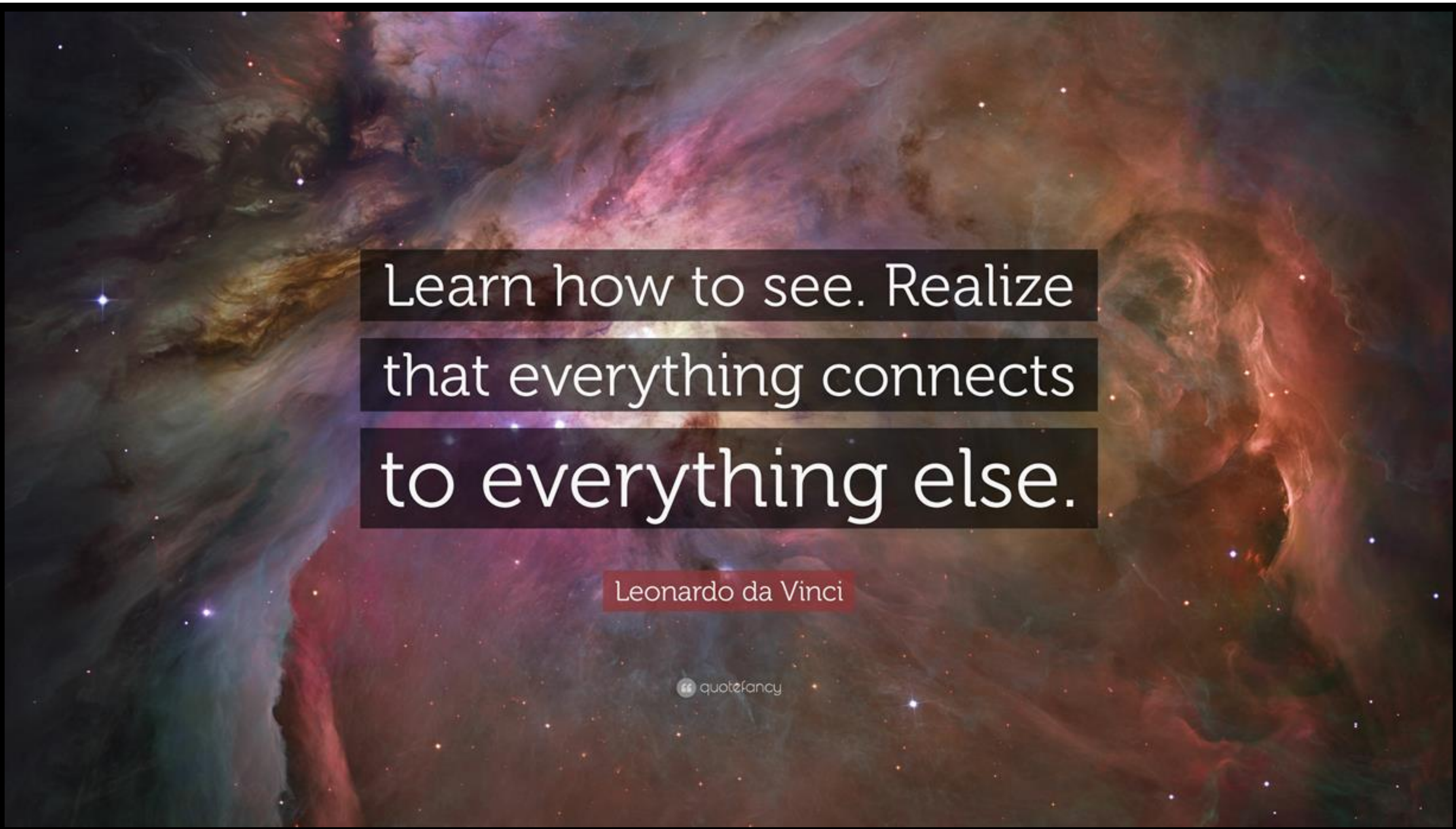
Our challenge: Train the A.I. to see the world in the ways that we do and to act on that insight!

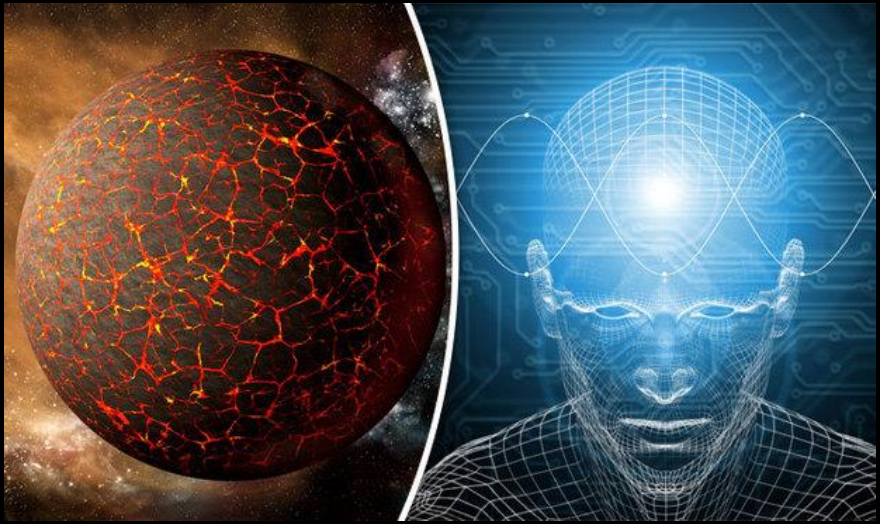


Our challenge: Train the A.I. to see the world in the ways that we do and to act on that insight!



Our challenge: Train the A.I. to see the world in the ways that we do and to act on that insight!





Come for the Data. Stay for the Science!

Thank you!

Twitter: @KirkDBorne or Email: kirk.borne@gmail.com

Get slides here: http://www.kirkborne.net/NASA-AI2018